Journal of Actuarial, Finance and Risk Management (JAFRM)
e-ESSN:28303938
Vol. 3(1), July, 2024.
http://e-journal.president.ac.id/presunivojs/index.php/JAFRM/index

Life Table Prediction Using the Lee-Carter Model

Putu Zeni Candrika Kertayanda[1], Tiara Yulita[2,*], Dila Tirta Julianty[3], Annastasya Ginting[4], Mei Nalda Adelia Simanjuntak[5]

[1,2,3,4,5] *Program Studi Sains Aktuaria, Jurusan Sains, Institut Teknologi Sumatera, l. Terusan Ryacudu, Way Huwi, Kec. Jati Agung, Kabupaten Lampung Selatan, Lampung, 35365, Indonesia*

[*]*Corresponding author*: *tiara.yulita@at.itera.ac.id*

Receive: July, 2, 2024, Revised: July, 19, 2024, Accepted: July, 30, 2024

Abstract— Mortality is the state where all signs of life permanently, which can occur at any time after a person is born. Mortality data can be presented in the form of table called a mortality table, which is a table that provides an overview of the life of a population group starting from births at the same time and life of a population group that starts from birth at the same time and slowly decreases due to death. In actuarial studies, this table has a role as one of the important factors to determine the amount of premium costs and premium reserves, especially life insurance. Therefore, it is necessary for continuity in conducting research related to mortality tables. This research aims to forecast the death rate and the probability of death of the population in the future using the Lee-Carter model, which is a mortality forecasting model that combines a demographic model with a time series model. This mortality model shows that the logarithm of mortality rate is the sum of the parameters of the average general mortality rate by age and the multiplication of the trend parameter of mortality rate changes by age with the mortality index parameter. Forecasting process begins with estimating the parameters of the average mortality rate and the trend of mortality rate changes influenced by the mortality index parameter using the singular value decomposition (SVD) method. After that, the mortality index mortality index is forecasted using the ARIMA model and the results of this forecast are then reinserted into the Lee-Carter model to obtain death rate prediction. Based on the results of the mortality rate prediction, it can then be the prediction of the probability of death for the mortality table. The result of this research is a mortality table that contains predictions of the probability of death of the Indonesian population for both male and female gender using the Lee-Carter model from the year 2022 2026. Based on these results, it is concluded that the value of the probability of death for each year increases with increasing age.

Keywords— Forecasting, Lee-Carter Model, Life Table, Mortality.

I. INTRODUCTION

Mortality is a statistic that reflects the number of individuals in a population who die within a specific period [1]. Generally, mortality is presented in the form of a mortality table, also known as a life table, which represents mortality data structure in probability format. Additionally, mortality succinctly illustrates the concept that the gradual impact of mortality reducing population numbers can be explained through tabulation [2].

Mortality tables play a crucial role in demographic, epidemiological, and actuarial studies. In demography, mortality tables are often used for descriptive purposes to compare mortality rates across different ages, genders, races, times, and locations. In epidemiology, mortality tables are employed to identify risk factors associated with morbidity and mortality rates [3]. In actuarial studies, mortality tables are essential for determining premium costs and premium reserves, particularly for life insurance. Moreover, mortality rates can be used as a benchmark for assessing public health and welfare, as well as for guiding government policy and evaluating initiatives, such as health service development, schools, public facilities, and other necessary infrastructures aimed at reducing mortality rates. Countries with better development progress, such as improved health, education, and higher economic levels, generally exhibit lower mortality rates [4].

Mortality tables can be constructed using approaches from the laws of mortality, namely deterministic mortality laws and stochastic mortality laws. Deterministic mortality laws were developed earlier than stochastic methods and tend to treat death rates as fixed values at each point in time. Some of the most notable deterministic laws in history include those by De Moivre, Gompertz, Makeham, and Weibull [5]. On the other hand, stochastic

mortality laws do not disregard the uncertainty element in mortality, treating it as a random variable. One prominent model under stochastic mortality laws is the Lee-Carter model.

The Lee-Carter model is widely recognized for its simplicity and effectiveness in forecasting mortality rates. It uses only three parameters, which makes it computationally efficient and easy to interpret. This model has been successfully applied in various countries, demonstrating its robustness and versatility. However, it assumes homoskedasticity of errors, which is unrealistic as the variance of mortality rates often increases with age. Another limitation is its assumption that the sensitivity of log mortality rates at each age remains constant, neglecting potential age-time interactions. Extensions like the Poisson Lee-Carter model address some of these issues by not imposing homoskedasticity and ensuring the total actual deaths match the total expected deaths. Nevertheless, these extensions can be computationally complex and may not always converge. Furthermore, incorporating cohort effects, as in the age-period-cohort Lee-Carter model, can improve performance but adds complexity and potential identifiability issues. The two-step estimation process in the Lee-Carter model can also lead to inconsistent estimators, particularly when dealing with non-stationary time series. Despite these challenges, the Lee-Carter model remains a cornerstone in mortality forecasting due to its balance of simplicity and effectiveness [6].

The Lee-Carter model was first introduced by Ronald D. Lee and Lawrence Carter in 1992 in their article "Modeling and Forecasting U.S. Mortality Rates" [7]. The Lee-Carter model is a forecasting model that combines demographic modeling with time series statistical methods [1]. The initial application of the Lee-Carter model performed well in generating mortality data for the United States from 1933 to 1987. Since then, the Lee-Carter model has been widely applied in various countries to forecast population mortality rates.

Based on previous research, it was found that mortality rate forecasts using the Lee-Carter model could be used to estimate mortality probabilities in life tables [1]. The study titled "Forecasting Malaysian Mortality Rates Using the Lee-Carter Model with Fitting Period Variants" concluded that the Lee-Carter model remains valid for mortality data forecasting [8]. Furthermore, other studies have also reported the successful application of the Lee-Carter model for mortality rate forecasting in different countries and time periods, such as in Canada, Brazil, and Belgium [9].

In this study, the researchers will use the Lee-Carter model to forecast future mortality rates and probabilities for the Indonesian population. The data to be used comprises mortality probabilities for the Indonesian population by single age and gender from 1967 to 2021, sourced from the 2022 World Population Prospects report released by the United Nations, with forecasting age limited to up to 99 years. The objective of this study is to construct mortality tables containing forecasted mortality probabilities for the Indonesian population from 2022 to 2026 using the Lee-Carter model.

II. LITERATURE REVIEW

A. Singular Value Decomposition (SVD)

Singular value decomposition (SVD) is a matrix decomposition method that decomposes a matrix into three simpler matrices [10]. SVD factors an $m \times n$ matrix **A** into matrices $\boldsymbol{U}$, $\sum$, $\boldsymbol{V}$.

$$\boldsymbol{A} = \boldsymbol{U}\sum\boldsymbol{V^T}. \tag{1}$$

Explanation:

$\boldsymbol{U}$: an orthogonal matrix of dimensions m×m, referred to as the left singular vectors.

$\sum$: a matrix of dimensions $m \times n$ where the main diagonal elements are the singular values of matrix $\boldsymbol{A}$, and the other elements are 0.

$\boldsymbol{V}$: an orthogonal matrix of dimensions $n \times n$, referred to as the right singular vectors.

Here is the algorithm of singular value decomposition method for matrix $\boldsymbol{A}$ of size $m \times n$ with rank k:

1. Left Singular Vector
   a. Form the matrix $\boldsymbol{AA^T}$.
   b. Calculate the eigenvalues of $\lambda_1, \lambda_2, \cdots, \lambda_n$ of $\boldsymbol{AA^T}$.
   c. Calculate the eigenvectors, namely $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_m}$ which correspond to eigenvalues, namely $\lambda_1, \lambda_2, \cdots, \lambda_n$ of $\boldsymbol{AA^T}$.

$$\begin{aligned} \hat{u}_m &= \frac{u_m}{\|u_m\|}, \\ &= \frac{u_m}{\sqrt{u_{1m}^2+u_{2m}^2+\cdots+u_{nm}^2}}. \end{aligned}$$

   d. Normalization of eigenvectors, namely $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_m}$ by dividing each vector by its length (magnitude).
   e. The shape of the matrix $\boldsymbol{U}$.

$$\boldsymbol{U} = [u_1 \quad u_2 \quad \ldots \quad u_k \quad | \quad u_{k+1} \quad \ldots \quad u_m].$$

2. Singular Values
   a. Form the matrix $\boldsymbol{A^T A}$.
   b. Calculating the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ of $\boldsymbol{A^T A}$.

c. Calculate the singular values of matrix $\boldsymbol{A^T A}$ $\sigma_1, \sigma_2, \cdots, \sigma_n$ with equation:

$$\sigma_n = \sqrt{\lambda_n}.$$

d. The form of the matrix $\sum$, which is a matrix whose main diagonal contains the singular values of the matrix $\boldsymbol{A^T A}$ and other values 0 assuming $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

$$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & 0 & \ldots & 0 & \\ 0 & \sigma_2 & \ldots & 0 & 0_{k\times(n-k)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \ldots & \sigma_k & \\ \hline & 0_{(m-k)\times k} & & & 0_{(m-k)x(n-k)} \end{array}\right].$$

3. Right Singular Vector

a. Calculate the eigenvectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_m}$ that correspond to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ of $\boldsymbol{A^T A}$.

b. Normalize the eigenvectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_m}$ by dividing each vector by its length (magnitude).

$$\hat{v}_m = \frac{v_m}{\|v_m\|}, = \frac{v_m}{\sqrt{v_{1m}^2 + v_{2m}^2 + \cdots + v_{nm}^2}}.$$

c. The matrix form $\boldsymbol{V}$.

$$\boldsymbol{V} = [v_1 \quad v_2 \quad \ldots \quad v_k \quad | \quad v_{k+1} \quad \ldots \quad v_m].$$

d. Transpose the matrix $\boldsymbol{V}$.

$$\boldsymbol{V^T} = \left[\begin{array}{c} V_1^T \\ V_2^T \\ \vdots \\ V_k^T \\ \hline V_{k+1}^T \\ \vdots \\ V_n^T \end{array}\right].$$

B. Lee-Carter Model

The Lee-Carter model is a model that was first introduced by Ronald D. Lee and Lawrence Carter in 1992 in the article Modeling and Forecasting the Time Series of U.S. Mortality [7]. This model is a mortality forecasting model that combines demographic models with time series statistical models [1]. The first application of the Lee-Carter model was said to perform well in producing the output data of the United States mortality rate from 1933 1987. Since then, the Lee-Carter model has been applied in many countries to forecast population mortality rates, namely Canada, Chile, Japan, Brazil, Austria, and Belgium. This model is widely used due to its simplicity, which only involves three parameters, but can forecast mortality rates well.

The basic concept of the Lee-Carter model is a stochastic model that suggests a log bilinear form for central death rates ($m_{x,t}$) for age x at time t [9]. The parameters contained in the Lee-Carter model, namely the a_x is parameter of the general average mortality rate by age x, b_x is trend parameter of the change in mortality rate by age x, k_t is mortality index parameter in year t. Here is the form Lee-Carter model equation:

$$m_{x,t} = e^{a_x + b_x k_t + \varepsilon_{x,t}}, \tag{2}$$

or

$$\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}. \tag{3}$$

where a_x, b_x, and k_t are the parameters to be estimated and $\varepsilon_{x.t}$ is a set of random errors. The parameters in Equation (3) are not unique because they are invariant to transformation [11]. Suppose, the vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{k}$ are one solution then for every c, $a - bc$, b, and $k + c$ are also solutions. In addition, if $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{k}$ are a solution then a, bc, $\frac{k}{c}$ is also a solution. Therefore, to get the solution unique from the parameter estimation in Equation (3), parameter constraints are made,

$$\sum_x b_x = 1, \tag{4}$$

and

$$\sum_t k_t = 0. \tag{5}$$

1) *Parameter Estimation* a_x: Based on the Lee-Carter model in Equation (3), the equation for the error value $\varepsilon_{x,t}$ can be obtained as follows [1]:

$$\varepsilon_{x,t} = \ln m_{x,t} - a_x - b_x k_t.$$

The value of a_x is estimated by minimizing the error value $\varepsilon_{x,t}$ or $\sum_{t=1}^{T}\varepsilon_{x,t}=0$ for a certain value of t so that it is obtained:

$$\begin{aligned}\sum_{t=1}^{T}\varepsilon_{x,t} &= \sum_{t=1}^{T}\left(\ln m_{x,t}-a_x-b_xk_t\right),\\ 0 &= \sum_{t=1}^{T}\left(\ln m_{x,t}-a_x-b_xk_t\right),\\ &= \sum_{t=1}^{T}\left(\ln m_{x,t}\right)-\sum_{t=1}^{T}(a_x)-\sum_{t=1}^{T}(b_xk_t),\\ &= \sum_{t=1}^{T}\left(\ln m_{x,t}\right)-Ta_x-b_x\sum_{t=1}^{T}(k_t).\end{aligned}$$

Thus,

$$Ta_x=\sum_{t=1}^{T}\left(\ln m_{x,t}\right)-b_x\sum_{t=1}^{T}(k_t),$$

or

$$a_x=\frac{\sum_{t=1}^{T}\left(\ln m_{x,t}\right)-b_x\sum_{t=1}^{T}(k_t)}{T} \tag{6}$$

Parameter estimation a_x as the average general mortality rate by age x obtained by using parameter constraints according to Equation (5) so that:

$$\begin{aligned}\hat{a}_x &= \frac{\sum_{t=1}^{T}(\ln m_{x,t})-b_x\cdot 0}{T},\\ &= \frac{\sum_{t=1}^{T}(\ln m_{x,t})}{T}.\end{aligned} \tag{7}$$

Description:
$m_{x,t}$: central death rate for age x at time t.
T : number of years.

2) Parameter Estimation of b_x and k_t: Parameter estimates of b_x and k_t are calculated using singular value decomposition (SVD) with the following steps [1]:

1. Form a matrix $\boldsymbol{Z_{x,t}}$ of size $x \times t$ to estimate the parameters b_x and k_t.

$$\boldsymbol{Z_{x,t}}=\begin{bmatrix} z_{0,0} & z_{0,2} & \cdots & z_{0,t}\\ z_{1,1} & z_{1,2} & \cdots & z_{1,t}\\ \vdots & \vdots & \ddots & \vdots\\ z_{x,1} & z_{x,2} & \cdots & z_{x,t}\end{bmatrix}.$$

The matrix $\boldsymbol{Z_{x,t}}$ is b_xk_t with b_xk_t formed based on Equation (2) whose error value is assumed to be zero so that it is obtained:

$$b_xk_t=\ln m_{x,t}-a_x.$$

Thus, the matrix $\boldsymbol{Z_{x,t}}$ is obtained as follows:

$$\boldsymbol{Z_{x,t}}=\begin{bmatrix} \ln m_{0,1}-a_0 & \ln m_{0,2}-a_0 & \cdots & \ln m_{0,t}-a_0\\ \ln m_{1,1}-a_1 & \ln m_{1,2}-a_1 & \cdots & \ln m_{1,t}-a_1\\ \vdots & \vdots & \ddots & \vdots\\ \ln m_{x,1}-a_x & \ln m_{x,2}-a_x & \cdots & \ln m_{x,t}-a_x\end{bmatrix} \tag{8}$$

2. The matrix $\boldsymbol{Z_{x,t}}$ is decomposed using SVD.

$$\boldsymbol{Z_{x,t}}=\boldsymbol{U\Sigma V^T}. \tag{9}$$

Description:
$\boldsymbol{U}$: left singular vector.
$\boldsymbol{\Sigma}$: singular values.
$\boldsymbol{V}$: right singular vector.

3. The estimation of b_x is obtained from the first column of matrix $\boldsymbol{U}$, i.e:

$$\hat{b}_x = (u_{1,1},\ u_{2,1},\ \cdots,\ u_{x,1})^T. \tag{10}$$

Based on the parameter constraints in Equation (6) then:

$$\hat{b}_x = \frac{1}{\sum_x u_{x,1}}(u_{1,1},\ u_{2,1},\ \cdots,\ u_{x,1})^T. \tag{11}$$

where $\hat{b}_x$ is estimation of trend parameter of the change in mortality rate by age x.

4. The estimation of k_t is obtained from the multiplication of the first singular value with column first matrix $\boldsymbol{V}$, that is:

$$\hat{k}_t = \Sigma_1(v_{1,1},\ v_{2,1},\ \cdots,\ v_{t,1}). \tag{12}$$

Based on the parameter constraints in Equation (6) then:

$$\hat{k}_t = \sum_x(u_{x,1})\cdot\Sigma_1\cdot(v_{1,1},\ v_{2,1},\ \cdots,\ v_{t,1}). \tag{13}$$

where $\hat{k}_t$ is estimation of mortality index parameter in year t.

3) Estimated Probability of Death: Based on the calculation of the estimated death rate ($m_{x,t}$) using the Lee-Carter model, the estimated value of the probability of death ($q_{x,t}$) can also be calculated. The probability of death can be obtained by finding the relationship between the central death rates of a person aged x in year t and the probability of death of a person aged x in year t. This relationship will be sought using linear interpolation. It is known that the equation for a person's death rate is:

$$m_x = \frac{d_x}{L_x},$$

with L_x is $L_x = \int_0^1 l_{x+t}\, dt$ and d_x is $d_x = q_x l_x$.
In linear interpolation, assume the linear form of l_{x+t} for $0 \le t \le 1$ as $a + bt$. Thus, for $t = 0$ then $l_x = a$ and for $t = 1$ then $l_{x+1} = a + b$ thus obtained [12]:

$$\begin{aligned} b &= l_{x+1} - a, \\ &= l_{x+1} - l_x. \end{aligned} \quad (14)$$

It is known that $-d_x = l_{x+1} - l_x$, so that Equation (14) becomes:

$$b = -d_x \quad (15)$$

From Equation (15), the following equation can be formed:

$$\begin{aligned} l_{x+t} &= a + bt, \\ &= l_x + (-d_x)t, \\ &= l_x - t \cdot d_x, \end{aligned} \quad (16)$$

Thus, the value L_x becomes:

$$\begin{aligned} L_x &= \int_0^1 (l_x - t \cdot d_x)\, dt, \\ &= \int_0^1 l_x\, dt - \int_0^1 d_x \cdot t\, dt, \\ &= t l_x|_0^1 - \tfrac{1}{2} d_x \cdot t^2|_0^1, \\ &= (l_x - 0) - \left(\tfrac{1}{2} d_x \cdot 1^2 - 0\right), \\ &= l_x - \tfrac{1}{2} d_x. \end{aligned} \quad (17)$$

Based on Equation (17), the equation for the central death rates of person aged x in Equation () becomes:

$$\begin{aligned} m_x &= \frac{d_x}{l_x - \frac{1}{2} d_x}, \\ &= \frac{l_x q_x}{l_x - \frac{1}{2} l_x q_x}. \end{aligned} \quad (19)$$

From Equation (19), the relationship between the death rate (m_x) and the probability of death (q_x) can be obtained, namely:

$$\begin{aligned} m_x &= \frac{l_x q_x}{\frac{2l_x - l_x q_x}{2}}, \\ &= \frac{2 l_x q_x}{2l_x - l_x q_x}, \\ &= \frac{2 l_x q_x}{l_x (2 - q_x)}, \\ &= \frac{2 q_x}{2 - q_x}. \end{aligned} \quad (20)$$

or

$$q_x = \frac{2 m_x}{2 + m_x}. \quad (21)$$

Thus, the equation to calculate the estimated value of the probability of death of a person aged x in year t ($q_{x,t}$) is as follows:

$$\hat{q}_{x,t} = \frac{2 \hat{m}_{x,t}}{2 + \hat{m}_{x,t}} \quad (22)$$

Description:
$\hat{m}_{x,t}$: estimated death rate of a person aged x in year t.

C. Autoregressive Moving Average

Autoregressive Integrated Moving Average (ARIMA) is one of the forecasting methods introduced by George E. P. Box and Gwilym M. Jenkins. This model Autoregressive (AR), Moving Average (MA), and ARMA models, which are a combination of AR and MA processes, are used for forecasting with non-stationary data. In addition, these models are also generated using the Box-Jenkins method [13].

The ARIMA model is an approach to building accurate forecasting models for the short term without considering independent variables, but only based on time series data. The general form is ARIMA (p, d, q), where p states the order of the autoregressive (AR) element, d is the order of the integrated (I) element, and q of moving average (MA) order [14]. The AR model reflects that the value of the dependent variable is influenced by the value of the variable itself at the previous time, while the MA model explains that the value of the dependent variable is influenced by the residual value (error) in the past [15]. The ARIMA model equation is as follows:

$$\Phi_p(B)D^d Z_t = \mu + \theta_q(B)a_t. \quad (23)$$

The above equation can be written using the B (backshift) operator as:

$$\left(1 - \emptyset_1 Y_{t-1} - \cdots - \emptyset_p Y_{t-p}\right)(1-B)^d Y_t = e_t + \theta_1 e_{t-1} + \cdots + \theta_p e_{p-q}$$

where Y_t is the value of the variable at time t.

Description:

Φ_p : the coefficient of the p^{th} autoregressive parameter.
θ_q : the coefficient of the q^{th} moving average parameter.
B : backshift operator.
D : differencing.
μ : constants.
a_t : the remainder at the t^{th} time
p : autoregressive degree.
d : the level of the differencing process.
q : degree moving average.

In the forecasting process using the ARIMA model, there are several stages, namely stationarity test, ARIMA model identification, and ARIMA model evaluation. The time series data used must be stationary, so a stationarity test is required, both a stationarity test on the average and a stationarity test in variance. The identification of the ARIMA model is seen based on the ACF and PACF plots. Meanwhile, for ARIMA model evaluation, the best model is selected based on Akaike's Information Criterion (AIC) approaches.

1) *Stationarity Test*: Time series data used in forecasting using the ARIMA method must be stationary, both in terms of mean and variance. Stationarity means that there is no change (constant) in the time series data. Stationary means that data fluctuations are around a constant mean value and the variance of these fluctuations remains constant over time [16].

   Data stationary test on variance can be done using the Box-Cox assumption test. The results of the Box-Cox assumption test using software assistance R Studio, it can be seen that based on the largest PPCC value, it is close to the value of nilai $\lambda = 1$ or not. If the largest PPCC value is not yet close to the value of nilai $\lambda = 1$ then it will be a transformation process is performed. The Box-Cox transformation or also known as the power transformation was introduced by Box and Cox in 1964. This transformation is not only useful for variance stationarity but also can often improve the approximation of the distribution to a normal distribution [17]. Box-Cox Transformation is the rank transformation on the response variable with λ is the estimated value transformation parameter and Zt is the response variable Z at time t [18]. This power transformation only applies to positive data. If some data values are zero or negative, a positive constant can be added to all data values to make them positive before they are transformed [19]. The following are Box-Cox transformation table for commonly used λ values:

Table 1. Box-Cox Transformation

λ Value	Transformation Shape
-1	$\frac{1}{Z_t}$
-0.5	$\frac{1}{\sqrt{Z_t}}$
0	$\ln Z_t$
0.5	$\sqrt{Z_t}$
1	Z_t

   Meanwhile, the process of testing the means is done using the Augmented Dickey Fuller test (ADF test). This stationarity test is based on the null hypothesis that the data is not stationary. Decision making using the ADF test is based on the p-value. The p-value is compared with the significance level value to determine whether or not a data is stationary. If the null hypothesis is accepted, which means that the data is not stationary, a differencing process needs to be carried out [20]. The differencing process is a calculation process by reducing the data value in a period with the data value in the previous data period [13]. The following is the hypothesis used in the ADF test:

   H_0 : Data is not stationary.

H_1 : Data is stationary.
Significance : $\alpha(5\%)$.
Rejection Area : reject H_0 if P-Value < $\alpha(0{,}05)$.

2) *ARIMA Model Identification*: Identification of the time series model can be done after the data is stationary against variance and to the mean. For p and q orders, it can be done by looking at the ACF and PACF plots, while for d orders, it can be seen from how many d orders are there. differencing [21]. The following are the general properties of ACF and PACF plots for ARIMA models [17]:

Table 2. ACF and PACF properties

Model	ACF	PACF
AR(p)	Goes to zero exponentially	Cut-off after p-lag
MA(q)	Cut-off after q-lag	Goes to zero exponentially
ARMA(p, q), $p > 0$ and $q > 0$	Goes to zero exponentially after lag-q	Goes to zero exponentially after lag-p

3) *Model Evaluation*: ARIMA model evaluation is determined based on Akaike's Information Criterion (AIC) approaches. The AIC approach was introduced by Akaike in 1973 to assess the quality of a good model with information criteria [22]. Model selection is done by minimizing the criterion function defined as follows:

$$AIC = -2\log(maximum\ likelihood) + 2k. \quad (24)$$

where k is the number of model parameters. The addition of $2k$ terms serves as a penalty function so that the selected model tends to be simpler, as well as for avoiding the selection of overly complex models with too many parameters. The main characteristic of AIC is its ability to evaluate the balance between the fit of the model to the data and the level of model complexity [23].

III. METHODOLOGY

A. Data Type and Source

The data used in this study is numerical or quantitative, consisting of numerical values. This data is sourced from the World Population Prospects 2022 report released by the United Nations. The study utilizes probability of death (q_x) from the UN (United Nation) Indonesia Life Table by single age up to 99 years for both males and females, covering the period from 1967 to 2021.

B. Variabels

This study involves two variables: year and probability of death (q_x) based on gender and single age. The time span used is 55 years, from 1967 to 2021, with single ages ranging from 0 to 99 years. These variables will be used to estimate the parameters of the Lee-Carter model, which include the parameter of the general average mortality rate by age x (a_x), the trend parameter of the change in mortality rate by age x (b_x), and the mortality index parameter in year t (k_t).

C. Research Procedure

The research procedure encompasses the steps used to collect data and solve the research problem. The procedure for this study is as follows:

1. Collect mortality probability data.
2. Calculate $m_{x,t}$ values using Equation (20), then compute $ln\, m_{x,t}$.
3. Estimate the parameter a_x by minimizing the error using Equation (7).
4. Form the matrix $\boldsymbol{Z_{x,t}}$ using Equation (8).
5. Estimate parameters b_x and k_t using singular value decomposition. The b_x estimate is obtained from the column of the matrix $\boldsymbol{U}$, and the k_t estimate is obtained by multiplying the first singular value with the first column of the matrix $\boldsymbol{V}$.
6. Test the stationarity of the k_t parameter data. If the data is not stationary with respect to variance, perform a transformation; if not stationary with respect to the mean, perform differencing.
7. Select the best ARIMA model based on the lowest AIC value.
8. Predict the mortality index parameter for year t (k_t) based on the best ARIMA model for future periods.
9. Substitute the estimated parameter values into the Lee-Carter Equation (2).
10. Forecast mortality probabilities using Equation (22).
11. Compile the Life Table.
12. Calculate the value of whole life insurance premium reserves using the prospective method with Equation ().
13. Draw conclusions.

D. Flowchart

A flowchart is a diagram that represents a process, system, or algorithm commonly used to document, plan, and illustrate a multi-step workflow. The following is the flowchart for this research:

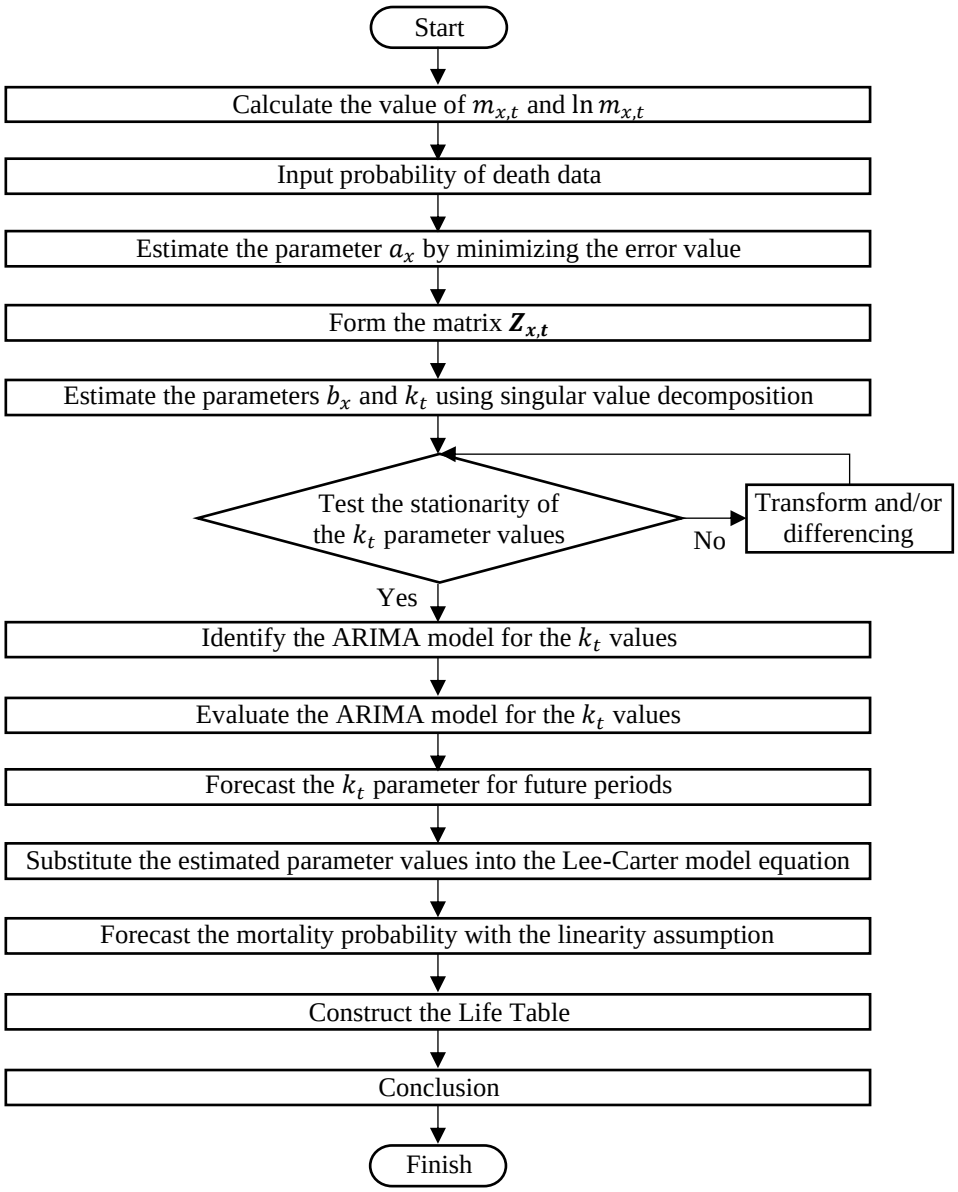

Figure 1. Flowchart

IV. RESULT AND DISCUSSION

A. Parameter Estimation a_x

The data used in this case study comprises probability of death (q_x) from the UN Indonesia Life Table, with ages $x = x_0, x_1, x_2, \dots, x_{99}$ and years $t = t_1, t_2, t_3, \dots, t_{55}$. Before estimating the parameter a_x, the first step is to calculate the values of $\ln(m_{x,t})$ using Equation (20):

$$\ln(m_{x,t}) = \ln\left(\frac{2q_{x,t}}{2 - q_{x,t}}\right)$$

A portion of the calculated $\ln(m_{x,t})$ values for males can be seen in Table 3 below:

Table 3. Result of $\ln(m_{x,t})$ calculation for male

Age (x)	**Year (t)**						
	1967	**1968**	**1969**	**⋯**	**2019**	**2020**	**2021**

0	-4.6315	-4.6721	-4.7208	⋯	-3.8152	-3.8487	-3.8821
1	-4.9588	-4.9974	-5.0448	⋯	-6.5232	-6.5712	-6.6189
2	-5.2456	-5.2825	-5.3287	⋯	-6.7163	-6.7593	-6.8009
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
97	-0.6897	-0.6884	-0.6904	⋯	-0.8025	-0.6709	-0.6129
98	-0.6383	-0.6369	-0.6389	⋯	-0.7454	-0.6205	-0.5653
99	-0.5913	-0.5899	-0.5918	⋯	-0.6915	-0.5726	-0.5199

Meanwhile, a portion of the calculated $\ln(m_{x,t})$ values for females can be seen in Table 4 below:

Table 4. Result of $\ln(m_{x,t})$ calculation for female

Age (x)	**Year (t)**						
	1967	**1968**	**1969**	⋯	**2019**	**2020**	**2021**
0	-4.6315	-4.6721	-4.7208	⋯	-3.8152	-3.8487	-3.8821
1	-4.9588	-4.9974	-5.0448	⋯	-6.5232	-6.5712	-6.6189
2	-5.2456	-5.2825	-5.3287	⋯	-6.7163	-6.7593	-6.8009
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
97	-0.6897	-0.6884	-0.6904	⋯	-0.8025	-0.6709	-0.6129
98	-0.6383	-0.6369	-0.6389	⋯	-0.7454	-0.6205	-0.5653
99	-0.5913	-0.5899	-0.5918	⋯	-0.6915	-0.5726	-0.5199

The next step is to estimate the parameter a_x by calculating the average mortality rate by age using Equation (7):

$$\hat{a}_x = \frac{\sum_{t=1}^{55}(\ln m_{x,t})}{55}$$

Table 5. Parameter estimation of $\hat{a}_x$

Age (x)	**Estimation of $\hat{a}_x$**	
	Male	**Female**
0	-2.8727	-3.0314
1	-5.0432	-5.1317
2	-5.3467	-5.4467
⋮	⋮	⋮
97	-0.7299	-0.8210
98	-0.6767	-0.7622
99	-0.6275	-0.7073

The estimated parameter $\hat{a}_x$ results in Table 5 can be visualized in Figure 2 and Figure 3. Based on the plot of $\hat{a}_x$ in Figure 2 and Figure 3, it can be observed that the estimated parameter $\hat{a}_x$ shows an increasing trend for both males and females. This indicates that in the younger age range, there is a tendency for lower average mortality rates, while in the older age range, there is an increase in the average mortality rate. However, it is found that individuals aged 0 years, or newborns, have a relatively high average mortality rate. Therefore, it can be concluded that there is a tendency for the average mortality rate to increase with age.

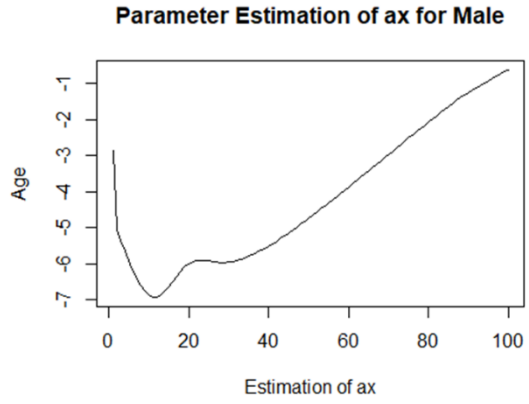

Figure 2. Parameter plot of $\hat{a}_x$ for male

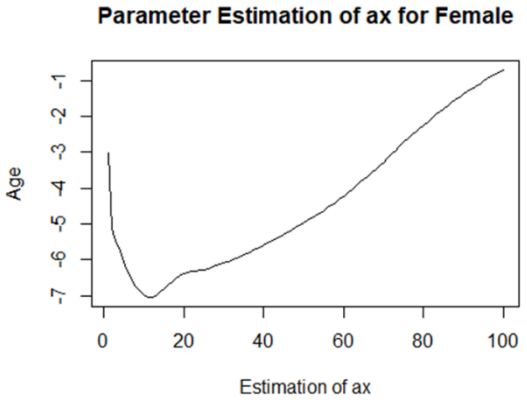

Figure 3. Parameter plot of $\hat{a}_x$ for female

B. Parameter Estimation of b_x and k_t

The estimation of parameters b_x and k_t is performed using singular value decomposition (SVD). The first step is to form the matrix $\boldsymbol{Z_{x,t}}$ based on Equation (8). Thus, the resulting $\boldsymbol{Z_{x,t}}$ matrix with dimensions 100×55 is as follows:

$$Z_{x,t} = \begin{bmatrix} Z_{0,0} & Z_{0,2} & \cdots & Z_{0,55} \\ Z_{1,1} & Z_{1,2} & \cdots & Z_{1,55} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{100,1} & Z_{100,2} & \cdots & Z_{100,55} \end{bmatrix}$$

$$Z_{x,t,male} = \begin{bmatrix} 0.91714 & 0.88949 & \cdots & -1.0095 \\ 1.53050 & 1.48381 & \cdots & -1.5757 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$Z_{x,t,female} = \begin{bmatrix} 0.94787 & 0.91766 & \cdots & -1.0578 \\ 1.56706 & 1.51626 & \cdots & -1.7041 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

The next step is to decompose the matrix $\boldsymbol{Z_{x,t}}$ using Equation (9). In this study, R-Studio software is utilized to decompose the matrix $\boldsymbol{Z_{x,t}}$ using SVD.

1) *Parameter Estimation of b_x*: The estimation of parameter b_x can be obtained using Equation (11). Below are partial results of the estimation of parameter b_x.

Table 6. Estimation parameter of $\hat{b}_x$

Age (x)	Estimation of $\hat{b}_x$	
	Male	**Female**
0	-2.8727	-3.0314
1	-5.0432	-5.1317
2	-5.3467	-5.4467
⋮	⋮	⋮
97	-0.7299	-0.8210
98	-0.6767	-0.7622
99	-0.6275	-0.7073

The estimated parameter $\hat{b}_x$ results in Table 6 can be visualized in the Figure 4 and Figure 5. The parameter $\hat{b}_x$ illustrates the sensitivity of the central mortality rate to changes in the mortality rate over time $\hat{k}_t$ at age x. From the illustrations in Figure 4 and Figure 5 above, it can be observed that the estimated values of parameter $\hat{b}_x$ for both genders, males and females, show a decreasing trend. For someone aged 0 years or a newborn, the value is relatively low and experiences a significant increase upon reaching 1 year of age. This phenomenon indicates that the pattern of the central mortality rate's influence on $\hat{k}_t$ tends to decrease with age.

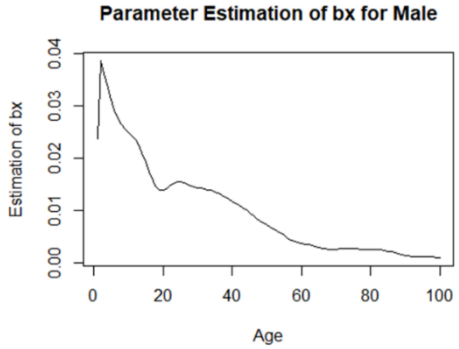

Figure 4. Parameter plot of $\hat{b}_x$ for male

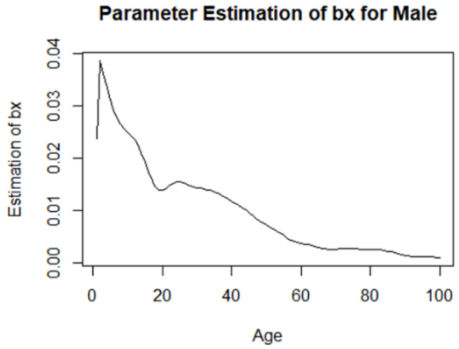

Figure 5. Parameter plot of $\hat{b}_x$ for female

2) *Parameter Estimation of k_t*: The estimation of parameter k_t can be obtained using Equation (13). Below are partial results of the estimation of parameter $\hat{k}_t$.

Table 7. Parameter estimation of $\hat{k}_t$

Year (t)	Estimation of $\hat{k}_t$	
	Male	**Female**
1967	0.023761	0.023042
1968	0.038737	0.038223
1969	0.036142	0.034945
⋮	⋮	⋮
2019	0.001072	0.000936
2020	0.000999	0.000789
2021	0.000000	0.000000

The estimated parameter $\hat{k}_t$ results in Table 7 can be visualized in the following plot:

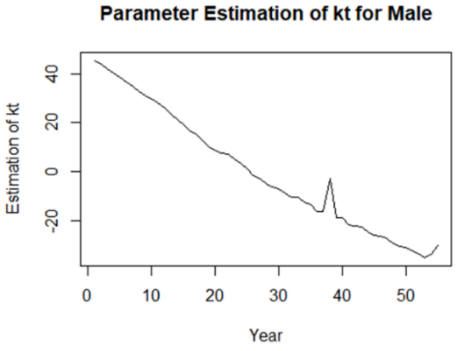

Figure 6. Parameter plot of $\hat{k}_t$ for male

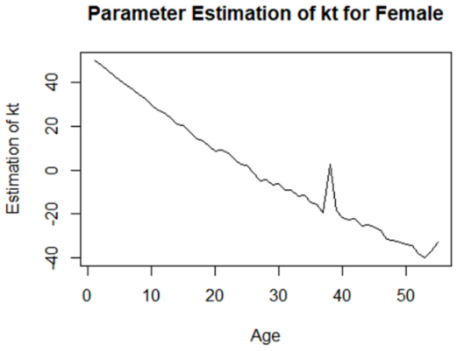

Figure 7. Parameter plot of $\hat{k}_t$ for female

The parameter $\hat{k}_t$ represents the mortality rate in period t. Figures 6 and 7 illustrate that in 1967, the mortality rate was quite high for both males and females. There was a significant increase in 2004, and in 2021, there was a slight increase compared to the previous year. Overall, the visual data indicate that the mortality rate parameter shows a decreasing trend annually from 1967 to 2021.

C. Parameter Prediction of $\hat{k}_t$ with ARIMA

The prediction of parameter $\hat{k}_t$ is obtained using the ARIMA method. The $\hat{k}_t$ parameter is predicted for the next five-year period. The ARIMA forecasting process begins with examining the stationarity of the $\hat{k}_t$ data, followed by time series model identification and model evaluation. Once the best model is selected, the next step is to forecast or predict for several future periods.

1) *Stationarity Test*: In this study, the Box-Cox test is used to examine the stationarity of the data with respect to variance. The data is considered stationary in variance if the maximum PPCC (Power of Principal Component Correlation) value is close to $\lambda = 1$. If the maximum PPCC value is not close to $\lambda = 1$, a transformation process will be performed. The following are the results of the Box-Cox test conducted using R-Studio software.

Table 8. Box-Cox test result of $\hat{k}_t$ for male and female

λ	PPCC	
	Male	**Female**
-2.0	0.4314109	0.3840666
-1.5	0.5169882	0.4527924
-1.0	0.6579098	0.5907800
-0.5	0.8304286	0.7938237
0.0	0.9488413	0.9428356
0.5	0.9840573	0.9869434
1.0	0.9745968	0.9804364
1.5	0.9502284	0.9571821
2.0	0.9230443	0.9294233

Based on Table 8, it can be seen that the highest PPCC values, 0.9840573 for males and 0.9869434 for females, are at $\lambda = 0.5$. This indicates that the data is not yet stationary with respect to variance. Therefore, according to the Box-Cox transformation criteria on Table 1, the data needs to be transformed using $\sqrt{Z_t}$. The following are the results of the Box-Cox test for the transformed data using R-Studio software:

Table 9. Box-Cox test result with transformation of $\hat{k}_t$ for male and female

λ	PPCC	
	Male	**Female**
-2.0	0.6579098	0.5907800
-1.5	0.7450354	0.6894723
-1.0	0.8304286	0.7938237
-0.5	0.9009240	0.8828485
0.0	0.9488413	0.9428356
0.5	0.9747741	0.9747622
1.0	0.9840573	0.9869434
1.5	0.9825143	0.9871380
2.0	0.9745968	0.9804364

In the Box-Cox transformation test, it was found that the highest PPCC value for males, 0.9840573, is at $\lambda = 1$, indicating that the data is stationary with respect to variance. However, for females, the highest PPCC value, 0.9871380, is at $\lambda = 1.5$, suggesting that the data is still not stationary with respect to variance. Therefore, the data for females needs to be further transformed using $Z_t\sqrt{Z_t}$. The following are the results of the Box-Cox test for the transformed data using R-Studio software:

Table 10. Box-Cox test results with data transformation of $\hat{k}_t$ for female

λ	**PPCC**
-2.0	0.4527924
-1.5	0.5482309
-1.0	0.6894723
-0.5	0.8414580
0.0	0.9428356
0.5	0.9827753
1.0	0.9871380
1.5	0.9754989
2.0	0.9571821

Based on the Box-Cox transformation test, it was found that the highest PPCC value, 0.9871380, is at $\lambda = 1$, indicating that the data is stationary with respect to variance.

Next, the stationarity of the data with respect to the mean will be tested. In this study, the Augmented Dickey-Fuller (ADF) test is used. The data tested is the previously transformed data. The hypotheses used for the ADF test on the $\hat{k}_t$ data for males and females are as follows:

H_0 : The data is not stationary.

H_1 : The data is stationary.

Decision-making in the ADF test is based on the p-value. The p-value is compared to the significance level to determine whether the data is stationary. The rejection region for this test is to reject the null hypothesis if the p-value is less than $\alpha = 5\%$. Based on the results, the p-value is 0.4220 for males and 0.1954 for females. According to the rejection region, the decision is not to reject the null hypothesis for both males and females. Therefore, it can be concluded that at a 5% significance level, the $\hat{k}_t$ data for males and females is not stationary with respect to the mean, so differencing is necessary.

After differencing, the data is re-tested using the ADF test. The results show a p-value of 0.01 for both males and females. According to the rejection region, the decision is to reject the null hypothesis. Therefore, it can be concluded that at a 5% significance level, the $\hat{k}_t$ data for males and females is stationary with respect to the mean. Thus, the data is ready to proceed to the next stage, which is time series model identification.

2) *Time Series Model Identification*: The time series model is identified based on the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots. Below are the ACF and PACF plots for the $\hat{k}_t$ data for males, generated using R-Studio software.

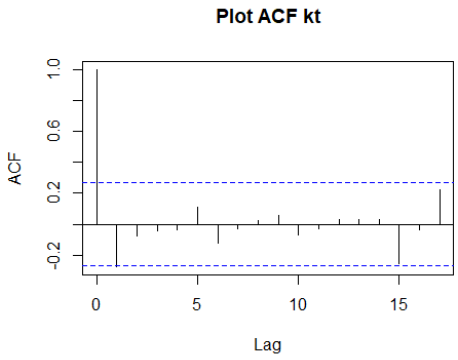

Figure 8. ACF plot of $\hat{k}_t$ for male

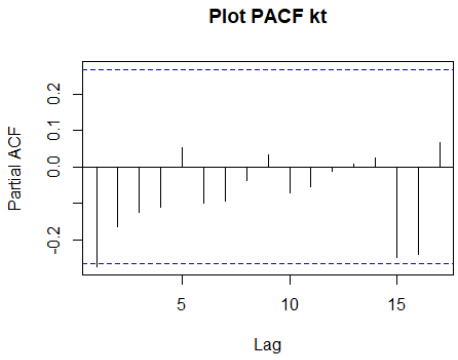

Figure 9. PACF plot of $\hat{k}_t$ for male

Based on the ACF and PACF plots above, it is observed that the ACF cuts off at lag 0 and 1, and the PACF cuts off at lag 1. Since there was 1 differencing step performed, the potential time series models could be:

1. ARIMA (1,1,0)
2. ARIMA (0,1,1)
3. ARIMA (0,1,0)
4. ARIMA (1,1,1)

For females, the ACF and PACF plots generated using R-Studio software are as follows:

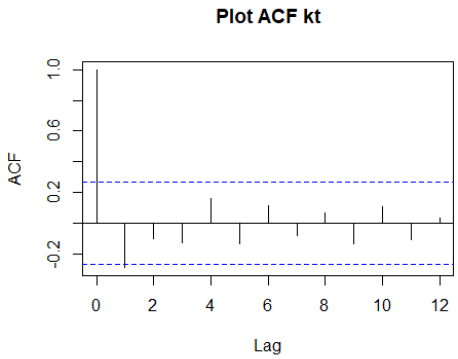

Figure 10. ACF plot of $\hat{k}_t$ for female

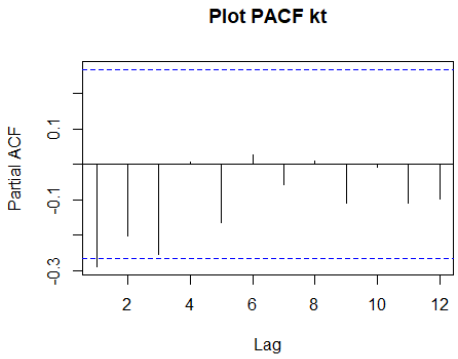

Figure 11. PACF plot of $\hat{k}_t$ for female

Based on the ACF and PACF plots above, it is observed that the ACF cuts off at lag 0 and 1, and the PACF cuts off at lag 1. Since there was 1 differencing step performed, the potential time series models could be:

1. ARIMA (1,1,0)
2. ARIMA (0,1,1)
3. ARIMA (0,1,0)
4. ARIMA (1,1,1)

3) *Model Evaluation*: Once the possible time series models are identified, the next step is to determine the best model by comparing the Akaike Information Criterion (AIC) values. The best model is the one with the smallest AIC value compared to the other models. The AIC values for each model, obtained using R-Studio software, are as follows.

Table 11. AIC Value for male and female

Model	AIC Value	
	Male	Female
ARIMA (1,1,0) with drift	268,36	308,31
ARIMA (0,1,1) with drift	264,42	300,20
ARIMA (0,1,0) with drift	277,38	318,20
ARIMA (1,1,1) with drift	266,30	302,12
ARIMA (1,1,0) without drift	286,24	320,00
ARIMA (0,1,1) without drift	286,69	320,89
ARIMA (1,1,1) without drift	287,81	321,95

Based on Table 11, it is observed that the smallest AIC values for both males and females are associated with Model number 2. Therefore, it can be concluded that the best model for forecasting the mortality index at year ttt $\hat{k}_t$ for both males and females is the ARIMA (0,1,1) model.

4) *Forecasting*: The final step in the time series analysis process is to forecast the mortality index $\hat{k}_t$ for several future periods. In this forecasting stage, the mortality index $\hat{k}_t$ will be predicted for the next five periods (2022-2026). Here are the forecasted $\hat{k}_t$ values.

Table 12. Prediction result of $\hat{k}_t$

Year	Prediction of $\hat{k}_t$	
	Male	Female
2022	−35.54365	−40.34549
2023	−37.00923	−41.96644
2024	−38.47482	−43.58739
2025	−39.94040	−45.20834
2026	−41.40599	−46.82929

Below are the graphs displaying the forecasted values of the mortality index parameter $\hat{k}_t$ for the years 2022-2026 using the ARIMA (0,1,1) model:

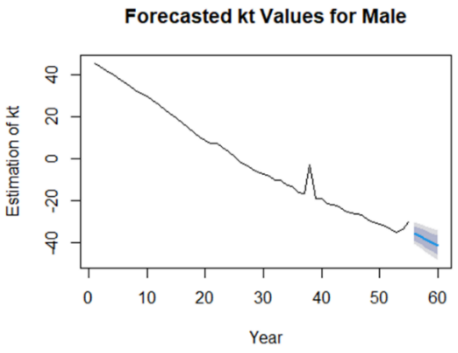

Figure 12. Plot of forecasted $\hat{k}_t$values for males

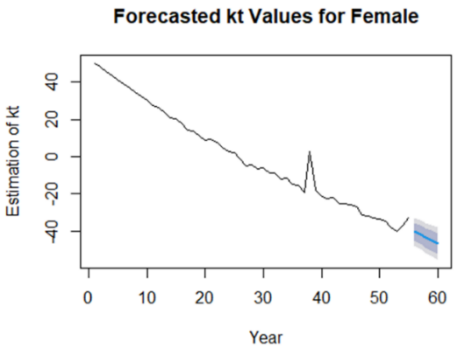

Figure 13. Plot of forecasted $\hat{k}_t$values for females

Based on Figures 12 and 13, it is evident that the forecasted values of $\hat{k}_t$ for 2022-2026 tend to decrease for both males and females.

D. Life Table

The mortality table constructed contains predicted probability of death for the years 2022-2026. The values of $\hat{q}_{x,t}$ are calculated using the relationship between central death rates and probability of death. Therefore, the first step is to determine the central death rates ($\hat{m}_{x,t}$). The predicted central death rates ($\hat{m}_{x,t}$) can be obtained by substituting the estimated parameters $\hat{a}_{x,t}$ from Table 5, $\hat{b}_x$ from Table 6, and the forecasted parameter $\hat{k}_t$ from Table 12 into the Lee-Carter model. A partial result of the predicted $\hat{m}_{x,t}$ for males is shown in the following table:

Table 13. Predicted central death rates for male

Age (x)	2022	2023	2024	2025	2026
0	0.02430	0.02347	0.02267	0.02189	0.02114
1	0.00163	0.00154	0.00145	0.00137	0.00130
2	0.00132	0.00125	0.00119	0.00112	0.00107
⋮	⋮	⋮	⋮	⋮	⋮
65	0.03233	0.03219	0.03206	0.03192	0.03179
66	0.03548	0.03534	0.03520	0.03506	0.03492
67	0.03888	0.03874	0.03859	0.03844	0.03829
⋮	⋮	⋮	⋮	⋮	⋮
97	0.46308	0.46232	0.46156	0.46080	0.46004
98	0.48927	0.48851	0.48774	0.48697	0.48621
99	0.51531	0.51456	0.51381	0.51305	0.51230

Meanwhile, a portion of the calculated $\hat{m}_{x,t}$ values for females can be seen in Table 14 below:

Table 14. Predicted central death rates for female

Age (x)	2022	2023	2024	2025	2026

0	0.01904	0.01835	0.01767	0.01703	0.01640
1	0.00126	0.00119	0.00112	0.00105	0.00099
2	0.00105	0.00099	0.00094	0.00089	0.00084
⋮	⋮	⋮	⋮	⋮	⋮
65	0.02018	0.01999	0.01980	0.01962	0.01943
66	0.02230	0.02210	0.02189	0.02169	0.02150
67	0.02469	0.02447	0.02425	0.02403	0.02382
⋮	⋮	⋮	⋮	⋮	⋮
97	0.42136	0.42063	0.41990	0.41917	0.41844
98	0.44935	0.44867	0.44799	0.44731	0.44663
99	0.47754	0.47693	0.47632	0.47571	0,47510

Next, the probability of death values ($\hat{q}_{x,t}$) are calculated using Equation (22). A portion of the predicted probability of death ($\hat{q}_{x,t}$) in the mortality table from 2022-2026 for males can be seen in the following table:

Table 15. Predicted probability of death for male

Age (x)	2022	2023	2024	2025	2026
0	0.02401	0.02320	0.02241	0.02165	0.02092
1	0.00163	0.00154	0.00145	0.00137	0.00130
2	0.00132	0.00125	0.00119	0.00112	0.00107
⋮	⋮	⋮	⋮	⋮	⋮
65	0.03181	0.03168	0.03155	0.03142	0.03129
66	0.03486	0.03472	0.03459	0.03445	0.03432
67	0.03814	0.03800	0.03786	0.03772	0.03758
⋮	⋮	⋮	⋮	⋮	⋮
97	0.37602	0.37552	0.37501	0.37451	0.37401
98	0.39311	0.39261	0.39212	0.39162	0.39113
99	0.40974	0.40926	0.40879	0.40831	0.40783

The predicted mortality probabilities for males in Table 15 can be visualized in Figure 14.
Based on the plot in Figure 14, it can be observed that the mortality probability values for Indonesian males increase with age for each year. Additionally, it can be seen that although the mortality probabilities from 2022 to 2026 tend to decrease, the plot of mortality probabilities for these years does not show significant differences. This is because the average decrease in mortality probabilities for Indonesian males from 2022 to 2026 is only 1%. Meanwhile, a portion of the calculated $\hat{q}_{x,t}$ values for females can be seen in Table 15.

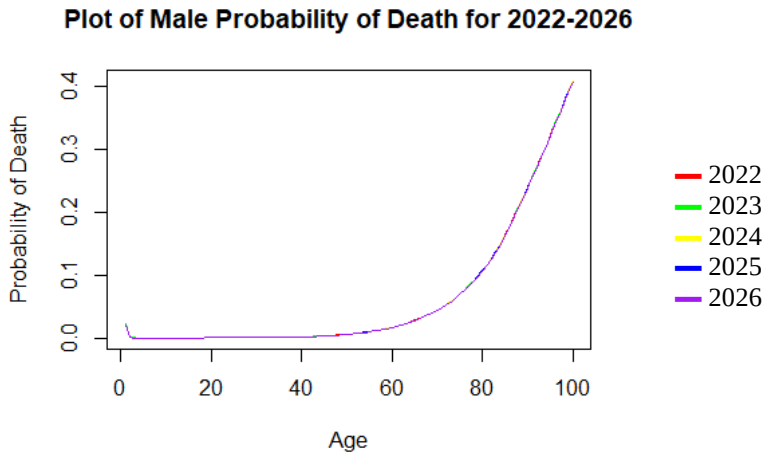

Figure 14. Plot of forecasted male probability of death for 2022-2026

Table 16. Predicted probability of death for female

Age (x)	2022	2023	2024	2025	2026
0	0.01886	0.01818	0.01752	0.01688	0.01627
1	0.00126	0.00119	0.00112	0.00105	0.00099
2	0.00105	0.00099	0.00094	0.00089	0.00084

⋮	⋮	⋮	⋮	⋮	⋮
65	0.01998	0.01979	0.01961	0.01943	0.01925
66	0.02205	0.02185	0.02166	0.02146	0.02127
67	0.02439	0.02417	0.02396	0.02375	0.02354
⋮	⋮	⋮	⋮	⋮	⋮
97	0.34804	0.34754	0.34704	0.34654	0.34604
98	0.36691	0.36646	0.36601	0.36555	0.36510
99	0.38550	0.38510	0.38470	0.38430	0.38391

The predicted mortality probabilities for males in Table 16 can be visualized in the following plot:

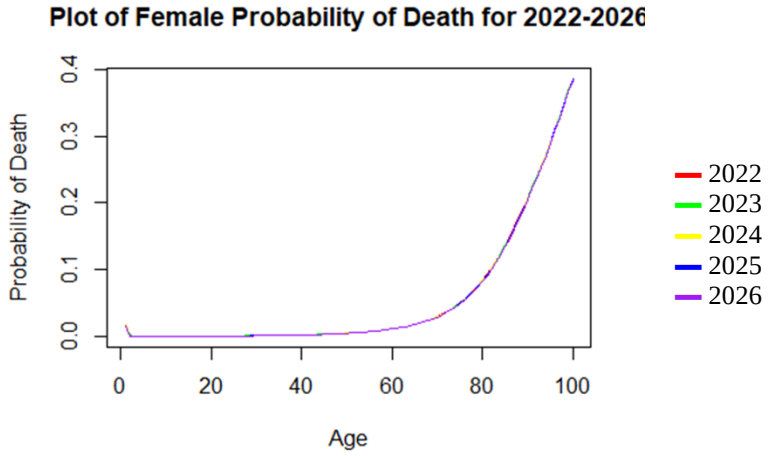

Figure 15. Plot of forecasted female probability of death for 2022-2026

Based on the plot in Figure 15, it can be observed that the mortality probability values for Indonesian females increase with age for each year. Additionally, it can be seen that although the mortality probabilities from 2022 to 2026 tend to decrease, the plot of mortality probabilities for these years does not show significant differences. This is because the average decrease in mortality probabilities for Indonesian females from 2022 to 2026 is only 2%.

E. Discussion

The Lee-Carter model, widely used for mortality forecasting, has several limitations and assumptions that should be considered for a comprehensive understanding of its results. One significant limitation is its reliance on a linear time trend in mortality improvement, which may not accurately capture sudden shifts or non-linear patterns due to unexpected events like pandemics or significant healthcare advancements. The model assumes that age-specific mortality improvements are proportional across all age groups, which might not hold true in reality, potentially introducing bias into the projections. Additionally, the model does not explicitly account for cohort effects, which can significantly impact mortality trends for specific generations. The accuracy of the model is also dependent on the quality and extent of historical data used for parameter estimation, making it less reliable in regions with incomplete or inconsistent data records. Notably, the data used in this study are projections, meaning the results may also be less accurate due to the inherent uncertainty in projection data. These limitations highlight the need for cautious interpretation of the model's forecasts, as they may not fully account for all variables influencing mortality rates.

To address these limitations and enhance the Lee-Carter model, several future developments and extensions can be considered. One approach is to incorporate cohort effects, as seen in extensions like the Renshaw-Haberman model, which can provide more accurate mortality forecasts by considering generational influences. Another potential improvement is the integration of non-linear time trends and machine learning techniques, which can better capture complex mortality patterns and abrupt changes. Additionally, using cause-specific mortality data can refine the model by accounting for different factors that affect mortality rates. Advances in computational power also open up possibilities for Bayesian methods, allowing for the incorporation of prior information and uncertainty estimation in forecasts. Moreover, extending the model to handle multi-population data can enhance its utility for comparative studies across different regions or demographic groups. These enhancements aim to provide more robust and adaptable tools for demographic forecasting, improving the reliability and applicability of the Lee-Carter model in various contexts.

V. CONCLUSION

A. Conclusion

Based on the results of the analysis and discussions conducted, it can be concluded that:

1. The forecasted life table for Indonesia using the Lee-Carter model for the years 2022-2026 tends to decrease, with the average percentage decrease is 1% for males and 2% for females.
2. The mortality probability values increase with age for each year, with an average percentage increase of 5%, for both males and females.

B. Recommendations

The recommendations provided for further research are as follows:

1. Utilizing the actual population data of Indonesia from 1967-2021 in predicting the Indonesian mortality table for 2022-2026.
2. Employing alternative models in constructing mortality tables, such as the Renshaw-Haberman model.

REFERENCES

[1] R. Unaeni, N. Setyahadewi and H. Perdana, "Peramalan Tingkat Kematian (Mortalita) Menggunakan Model Lee Carter," *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster),* vol. 6 (1), pp. 9-18, 2017.

[2] H. Utomo, "Perbandingan Tabel Mortalita Indonesia dan Tabel Mortalita CSO Menggunakan Uji Mann-Whitney dan Uji Kruskal-Wallis," *Syntax Literate: Jurnal Ilmiah Indonesia,* vol. 6 (3), p. 1210, 2021.

[3] R. K. Thomas, Concepts, Methods and Practical Applications in Applied Demography: An Introductory Textbook, Switzerland: Springer Cham, 2018.

[4] P. Omas Bulan Samosir, S. Rihlah Romdoniah and S. Israul Hasanah, Konsep dan Ukuran Mortalitas, Jakarta: Badan Kependudukan dan Keluarga Berencana Nasional, 2020.

[5] F. N. F. Sudding, "Pengaruh Asumsi Stokastik Mortalita Lee-Carter dan Asumsi Deterministik Gompertz Terhadap Nilai-Nilai Anuitas, Santunan, dan Premi," M.S. thesis, Institut Teknologi Bandung, 2014.

[6] L. Yang, Strength and Weakness of Lee-Carter Models, Netherlands: Maastricht University, 2019.

[7] Y. H. Umar and U. J. Chukwudi, "Modeling Mortality Rates Using Heligman-Pollard and Lee-Carter in Nigeria," *American Journal of Theoritical and Applied Statistics,* vol. 8 (6), pp. 221-239, 2019.

[8] N. S. M. Ibrahim, N. M. Lazam and S. N. Shair, " Forecasting Malaysian Mortality Rates Using the Lee-Carter Model with Fitting Period Variants," *Journal of Physics: Conference Series,* p. 9, 2021.

[9] I. Nursaadah, E. Puspita and R. Marwati, "Metode Peramalan Mortalita Menggunakan Metode Lee-Carter," *EurekaMatika,* vol. 3 (1), pp. 17-30, 2015.

[10] A. Sitanggang, A. D. Harahap, A. Karimullah, Y. A. Dewantara and C. Rozikin, "Sistem Rekomendasi Anime Menggunakan Metode Singular Value Decomposition (SVD) dan Cosine Similarity," *Jurnal Teknologi Informasi,* vol. 2(2), pp. 90-94, 2023.

[11] F. Girosi and G. King, "Understanding the Lee-Carter Mortality Forecasting Method," 2007.

[12] F. Dick London, Survival Models and Their Estimation Third Edition, ACTEX Publications, 1997.

[13] F. Fejriani, M. Hendrawansyah, L. Muharni, S. F. Handayani and Syaharuddin, "Forecasting Peningkatan Jumlah Penduduk Berdasarkan Jenis Kelamin Menggunakan Metode ARIMA," *Jurnal Kajian Geography,* vol. 8 (1), pp. 28-29, 2020.

[14] S. P. Fauzani and D. Rahmi, "Penerapan Metode ARIMA dalam Peramalan Harga Produksi Karet di Provinsi Riau," *Jurnal Teknologi dan Manajemen Industri Terapan,* vol. 2 (4), pp. 269-277, 2023.

[15] W. Rusyida and V. Pratama, "Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA," *SQUARE: Journal of Mathematics and Mathematics Education,* vol. 2 (1), p. 75, 2020.

[16] J. Purnama and A. Juliana, "Analisa Prediksi Indeks Harga Saham Gabungan Menggunakan Metode ARIMA," *Cakrawala Management Business Journal,* vol. 2 (2), p. 459, 2019.

[17] W. W. S. Wei, Time Series Analysis Univariate and Multivariate Methods Second Edition, United States: Pearson Addison Wesley, 2005.

[18] T. Yunita, "Peramalan Jumlah Penggunaan Kuota Internet Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA)," *JOMTA Journal of Mathematics: Theory and Applications,* vol. 1 (2), pp. 17-18, 2019.

[19] S. M. I Wayan Sumarjaya, Modul Analisis Deret Waktu, Bukit Jimbranan, 2016.

[20] N. A. Wijoyo, "Peramalan Nilai Tukar Rupiah Terhadap USD dengan Menggunakan Model GARCH," *Kajian Ekonomi Keuangan,* vol. 20 (2), p. 182, 2016.

[21] S. Makridakis, S. C. Wheelwright and R. J. Hyndman, Metode dan Aplikasi Peramalan Edisi Kedua, Jakarta: Erlangga, 1995.

[22] R. Agustini, N. Hajarisman and S. Sunendiari, "Kriteria Pemilihan Model Peramalan Terbaik Berdasarkan Kriteria Informasi," *Prosiding Statistika,* vol. 4 (1), pp. 57-64, 2018.

[23] J. D. Cryer and K.-S. Chan, Time Series Analysis With Applications in R, 2nd ed, New York: Springer, 2008.